MIS Quarterly

RESEARCH NOTE

THE CREATION OF SOCIAL VALUE: CAN AN ONLINE HEALTH COMMUNITY REDUCE RURAL–URBAN HEALTH DISPARITIES?[1]

Jie Mein Goh
Beedie School of Business, Simon Fraser University, 8888 University Drive,
Burnaby, BC V5A 1S6 CANADA {jmgoh@sfu.ca}

Guodong (Gordon) Gao and Ritu Agarwal
Robert H. Smith School of Business, University of Maryland,
College Park, MD 20742 U.S.A. {ggao@rhsmith.umd.edu} {ragarwal@rhsmith.umd.edu}

The striking growth of online communities in recent years has sparked significant interest in understanding and quantifying benefits of participation. While research has begun to document the economic outcomes associated with online communities, quantifying the social value created in these collectives has been largely overlooked. This study proposes that online health communities create social value by addressing rural–urban health disparities via improved health capabilities. Using a unique data set from a rare disease community, we provide one of the first empirical studies of social value creation. Our quantitative analysis using exponential random graph models reveals patterns of social support exchanged between users and the variations in these patterns based on users' location. We find that, overall, urban users are net suppliers of social support while rural participants are net recipients, suggesting that technology-mediated online health communities are able to alleviate rural–urban health disparities. This study advances extant understanding of value production in online collectives, and yields implications for policy.

Keywords: Healthcare, online communities, social value, disparities, social network

Introduction

The striking growth in online communities across a variety of domains and topic areas has been widely acknowledged in the scholarly and practitioner literatures (Faraj et al. 2015; Ma and Agarwal 2007; Manchanda et al. 2015; Nevo and Furneaux 2012; Sun et al. 2012; Wimmer and Lewis 2010). With advances in technology enabling richer discourse and exchange in individuals' interactions (Kane and Fichman 2009), there is, unsurprisingly, significant interest in understanding the nature of outcomes associated with these collectives and their value creation potential (Armstrong and Hagel 2000). To the extent that the interactions among users participating in the community are value producing, research on online communities has theorized about and empirically explored various ways in which users of online communities create economic value, such as the potential to lower a company's costs, generate sales, and foster innovation (e.g., Armstrong and Hagel 2000; Manchanda et al. 2015; Trusov et al. 2009). However, while economic value is doubtless important, a large number of online communities are not sponsored by a particular company, nor do they have direct business implications, raising the question of how else the benefits of online communities might be conceptualized.

[1]Andrew Burton-Jones was the accepting senior editor for this paper. Gerald Kane served as the associate editor.

In this study, we investigate value creation in the context of online communities focused on health. Online health communities facilitate sharing, dissemination, and creation of health information by patients at an unprecedented level, and represent one of the fastest growing types of virtual collectives. While the knowledge exchanged within these collectives has the potential to generate substantial *social* value for participants, this important effect of technology-mediated communities has received limited scholarly attention. Considering that individuals' health represents one of the most consequential spheres of activity for them, the setting provides a rich environment within which to explore questions related to social value.

We pose the question: Can an online health community enable participants to create social value by helping to alleviate regional health disparities between rural and urban users? The existence of health disparities arising from location is well documented, with substantial research indicating that rural residents experience greater health adversity due to lack of access to healthcare knowledge and resources (Institute of Medicine 2005; Lengerich et al. 2005; Martin et al. 2013). Online health communities plausibly provide an alternative forum that transcends geographic constraints, representing a supportive social networking resource to ameliorate health disparities. This assertion, however, remains understudied. Anchoring our expectation for such social value creation in the theoretical framework of health capabilities (Ruger 2010) that draws upon the pioneering work of Nobel laureate Amartya Sen's capability approach to human development (Sen 1999), we suggest that exchanges within the community can be instrumental in amplifying the health capabilities of rural users.

We test our assertion related to a reduction in regional health disparities with a unique data set from an online community focused on a rare disease. The "rarity" of the disease makes it challenging to find support in traditional ways such as through access to offline resources or encounters with similar others who are experiencing the disease. Thus, this online community represents an ideal setting to examine the value potential of social support among patients. In our data we are able to observe all interactions occurring within the community over a 44 month time period, the geographic location of participants in each interaction, together with a number of additional participant characteristics. Our empirical results, stable across multiple robustness tests, indicate there is a net surplus of social support from urban to rural users, thereby providing evidence for our claim of social value creation.

This study makes several contributions to the existing literature. To the best of our knowledge, it is the first to posit and empirically examine the role of online communities in reducing regional health disparities. This theoretical proposition extends extant understanding of the ways in which online communities are value producing. Methodologically, the study uses relatively novel estimation methods. Prior research on online communities, especially studies that examine drivers of sustainability and user engagement, have typically overlooked network characteristics in response to identification issues that arise when using traditional regression techniques. This approach is limited in that, to the degree that the essence of a community is in the social interactions that occur, it does not account for higher level dependencies such as homophily and reciprocity which have been well-established in social network literature as important influences on outcomes. Researchers have noted that these limitations can be addressed by using analysis techniques such as exponential random graph models (ERGM) (Ahern and Harford 2014; Cranmer and Desmarais 2011; Faraj and Johnson 2011; Jackson et al. 2012; Lusher et al. 2013; Wimmer and Lewis 2010). We analyzed network interaction patterns in our data to quantify social support exchanges using ERGM to explicitly include higher order network characteristics when establishing the social value creation potential of the online community.

Conceptual Background

Online Communities and Value Creation

In recent years, online community research has increasingly begun to emphasize value creation and community outcomes. One stream of work focuses on the revenue potential of online communities for firms (Armstrong and Hagel 2000; Manchanda et al. 2015). For instance, Manchanda et al. (2015) explore the economic value created by customer engagement in firm-sponsored online communities, and Huang et al. (2014) study online customer behavior in the context of generation of ideas for Dell Computers. A second stream links the benefits of online community engagement to performance outcomes. Nevo and Furneaux (2012) find that knowledge access, trust, and bridging ties are related to team performance, while Porter and Donthu's (2008) analysis suggests that firms' efforts to provide high quality content and promote community member embeddedness have positive effects on individuals' beliefs about the sponsor.

Prior work on the value of online communities has been largely motivated by the benefits they generate for firms, such as customer engagement and sales. However economic returns reflect only one dimension of the value potential of online communities. Because online communities enable the exchange of knowledge and support (Faraj et al. 2011), they hold considerable promise for generating social value for participants on the platform.[2] With increasing interest in the notion of social responsibility, and coupled with the growing number of firm-sponsored online communities, the question of whether online communities generate quantifiable social value is significant. Establishing the social value of online communities could help firms that create and sponsor online communities generate a positive impact on society, fulfilling, in part, their corporate social responsibilities and providing a mechanism for enhancing reputation and trust. However, although there is substantial anecdotal evidence to support the *individual* benefits that online communities confer on participants (e.g., Wasko and Faraj 2005), we do not find large-scale empirical studies that demonstrate the production of social value in online communities.

The Health Disadvantages of Location

Our focus is on the quantification of social value as reflected in the extent to which an online health community is able to alleviate health disparities between rural and urban users. The term *health disparity* connotes a situation in which inequitable medical outcomes are observed as a result of some "social, economic or environmental disadvantage" (Koh et al. 2011). One key source of disparity is geographical location that can create a significant handicap for rural residents (Rutten et al. 2012). In spite of the fact that in certain cases rural populations may be endowed with a comparative advantage with respect to their urban counterparts (such as less industrial pollution, lower stress, and healthier lifestyles), a large number of studies show that rural communities experience greater health adversity, including higher rates of obesity, cancer, heart disease, diabetes, and injury-related deaths (Institute of Medicine 2005; Martin et al. 2013).

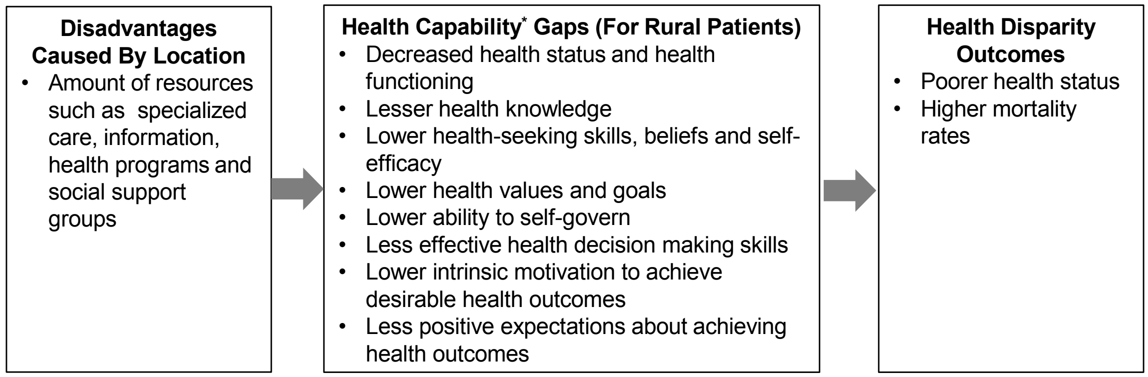
While many factors have been implicated in the existence of health disparities such as overt discrimination, socioeconomic status, and education (Adler and Rehkopf 2008), a central driver of health disparities between rural and urban areas is unequal access to health resources and information (Baldwin et al. 2004; United Health 2011). With 50 percent of the world's population living in rural areas, the difficulty in attracting and retaining medical professionals in these areas has consistently been identified as a global challenge and remains a major concern for world leaders (World Health Organization 2010). In the United States, only 11 percent of the total pool of primary care physicians practice in rural areas. The problem is particularly acute in specialized medical fields, with rural areas facing a chronic shortage of specialty physicians (Gamm et al. 2010; O'Grady et al. 2002; Schur and Franco 1999; Ziller 2014). Among rural physicians, 51 percent of the primary care physicians refer their patients to a distant city or town more than 20 miles from their office, and 47 percent of physicians believe that it is at least somewhat difficult for their patients to obtain specialty services locally (United Health 2011).

In addition to the paucity of qualified professionals, rural patients are further disadvantaged with respect to access to support groups. Previous work finds that access to and participation in support groups generate beneficial health impacts even in resource-poor environments (Nobles and Frankenburg 2009). Such patient networks can also serve the uninsured as a key source of medical guidance. However, geographic distance makes it difficult for rural patients to form such support groups, especially in the case of rare diseases. In rural areas, low population densities imply that the likelihood of finding individuals suffering from a certain disease is smaller, thereby limiting individuals' ability to obtain and provide information locally.

Limited access to health professionals and support groups creates significant disadvantages for rural patients. To conceptualize the nature of this handicap, we adopt a capabilities perspective on disparities (Ruger 2010), which draws on Sen's (2002) pioneering work on health equity to propose a health capabilities model (HCM). Sen's principal argument related to health equity emphasizes the importance of human capability to function—what people are effectively *able* to do and be (Sen 2002). HCM builds on this fundamental principle of capability to suggest that health outcomes are directly related to health agency—what individuals are *able* to do. The HCM model identifies eight internal factors (see Figure 1) that collectively constitute health capabilities. The disadvantages arising from location that rural patients experience result in a health capabilities gap, one of the fundamental causes of disparities in health outcomes. To address inequalities then, policy makers should be concerned with reducing inequities in health capabilities among individuals and groups rather than simply distributing resources equally among them. As Ruger (2010, p. 42) notes, "Conceptually, health capability enables us to understand the conditions that facilitate and barriers that impede health and the ability to make health choices."

The presence of health capability gaps among rural patients is documented in the literature. For instance, rural patients

[2]Notable examples illustrating such social value include the online communities on President Obama's campaign website and the U.S. Navy's "Navy for Moms," a community created for mothers with children in the Navy to facilitate member support and engagement.



*See Ruger (2010) for the definitions of each health capability.

Figure 1. Framework for Rural–Urban Health Disparities (adapted from Ruger 2010)

experience health capability gaps in the form of health knowledge; Ricketts' (2000) study identified a paucity of information in the areas of coping, treatment and side effects, and home care as critical unmet needs of rural patients. The reduced information availability also leads to lower awareness of health issues and, in turn, poorer health outcomes (Berkman et al. 2011). Studies on health literacy, or the degree to which individuals have the capacity to understand and process information necessary to make informed health decisions, underscore the fact that rural individuals tend to have lower health literacy, likely caused by limited interactions with both healthcare professionals and peer patients. Collectively, limited availability of health experts, lower levels of literacy in general and health literacy in particular, and a smaller pool of geographically proximate individuals bearing the burden of a specific disease all contribute to the lower health information resources of rural residents.

Enhancing the Health Capabilities of Rural Users

Given that rural users are disadvantaged relative to urban users, how might the interaction in an online health community address such disparities? We suggest that online communities reduce the health capabilities gap experienced by rural patients by enabling the exchange of *social support* (Figure 2). As with other communities structured around shared interests and goals, a significant benefit of participation in online health collectives is the wealth of knowledge and information possessed by participants (Franke and Shah 2003). In these forums, information and knowledge about the disease condition, treatments, treatment alternatives, and other disease-related advice (Bunde et al. 2006; Graydon et al. 1997; Wellman et al. 1996) can be a valuable resource for community members. Indeed, there is significant evidence of the positive benefits of such interactions, generally referred to as *social support* (House 2002) in improving patients' ability to cope with disease, reduce anxiety, and alleviate mood disturbances (Rutten et al. 2012; Viswanath 2006).

Online social support has the potential to enhance health capabilities in at least two ways. First, since rural patients tend to be disadvantaged in access to health information, online communities provide a platform for them to ask questions and learn from urban patients about the nature of the disease and treatment options. Second, the online community provides emotional support similar to that available through offline social support groups (e.g., Alcoholics Anonymous), which helps to maintain a positive outlook and better manage their disease conditions. In other words, both enhanced access to information and a heightened motivation to achieve desirable outcomes have the potential to enhance the health capabilities of rural users.

To summarize, we have suggested that social support exchanged in an online health community is positively associated with a reduction in health disparities caused by location via its ability to reduce the health capabilities gap of rural community members. To the degree that community interaction has a relatively *more* positive effect for rural patients, we can reasonably conclude that online communities generate social value by reducing rural–urban health disparities.

We next address the question of why online health communities are likely to be more efficacious in enhancing the health capabilities of rural users as compared with their urban counterparts. As discussed in the literature on regional disparities

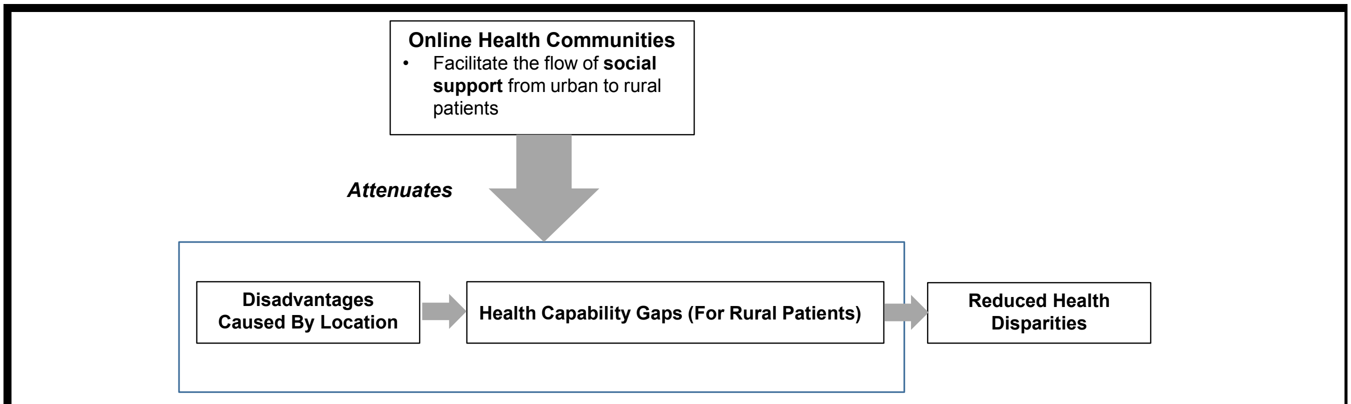


Figure 2. A Conceptual Model of Social Value Creation in Online Health Communities

(Lipsky and Glasser 2011), urban residents generally have greater access to needed informational resources through offline channels. This is especially true for rare diseases, where the dispersion of patients in rural areas makes access even more challenging. Urban patients' superior knowledge endowments allow them to serve as a source of expertise, while rural patients may lack the necessary information to contribute meaningfully to the online discourse. Thus, we would expect urban patients to have an overall greater *ability* to provide support to rural patients.

However, the ability to contribute to community discourse is necessary, but not a sufficient condition to enable the exchange of social support. Community participants must also be motivated to contribute. There is a sizable body of knowledge on the key motivations for user contribution and participation in online communities (e.g. Bateman et al. 2011; Ray et al. 2014; Tsai and Bagozzi 2014). Patients select into participating in an online health community volitionally, driven primarily by interest in the community discourse (Preece 2000). As they engage with others in the community, they develop a shared identity, commonly constructed around the specific disease that is the community's focus.[3] A common identity and shared goals of learning about the disease and managing adversity motivate *all* participants to help and support one another (Johnson and Ambrose 2006; Ren et al. 2007).

Therefore, while both rural and urban patients may have personal experience with the medical condition, the capacity to provide support is also critically dependent on geographic location. Since urban patients tend to have greater medical and information access compared to rural patients, their contributions of support are likely to be higher than those of patients living in rural areas. Based on this logic, we theorize the existence of a greater inflow of social support from urban patients to rural patients. It could alternatively be argued that urban patients, who receive greater support from other sources, might be less active in the online community. It remains an empirical question whether there is a net surplus of social support from urban to rural patients, which we investigate.

[3]To illustrate, breast cancer communities such as Breastcancer.org attract those afflicted with the medical condition.

Methods

Research Approach

We use a novel data set constructed from an online health community founded with the goal to facilitate the sharing of information that can help improve the lives of patients suffering from a specific rare disease. Two unique features make this disease an ideal research setting for studying the online community's role in reducing regional health disparities via the exchange of social support. First, the disease is caused by the degeneration of the nerve system and is unrelated to a patient's socioeconomic status. Although the disease has no known cure, particular therapies and care can help patients prolong survival. Thus, knowledge about various aspects of disease management and treatment options is critical for patients. Second, this disease is classified as "rare" based on the U.S. Rare Diseases Act of 2002 (i.e., the number of afflicted individuals within the United States is less than 200,000). Because of the low prevalence, individuals with rare diseases face challenges in connecting with others who have the same condition (Edwards 2013), and especially so in rural areas where population density and, therefore, disease incidence, is significantly lower. Collectively, these characteristics of the study setting create a relatively pure context for investigating social value creation.

The platform on which the focal community is constructed has a number of features, including a discussion forum for members to engage with each other, the capability to record their disease condition, symptoms, medications, and treatments in their profiles, and a search feature allowing them to view other patients' profiles. Other major functions available include the ability to post a message, search for relevant topics within the forum, subscribe to new replies for a particular topic, and tag topics. Because this online community is specialized to only focus on one disease (in contrast to a general purpose "wellness" community), all interactions within the platform center around this disease. This setting not only provides rich and complete data at the individual patient and network exchange levels, it also offers a contained and circumscribed context to investigate regional patterns of support exchange.

Sample and Data

We collected detailed data on all active users from this community for a period of 44 months from October 2005 through June 2009. We matched city and state information contained in users' profiles to the Area Resource File (ARF) database obtained from the United States Health Resources and Services Administration to classify the patients' location into urban or rural categories (based on a combination of Federal Information Processing and Metropolitan/Micropolitan Statistical Areas codes found in the ARF).[4] Given our interest in understanding the nature of interactions among rural and urban patients within the disease community in the United States, we removed non-patient members (e.g., caregivers), patients who did not participate in the forum, patients with incomplete data, and patients from other countries. This yielded a final dataset of 638 patients, among which there were 111 rural patients and 527 urban patients.

Our data contains a rich set of variables for each patient, including demographic, disease-related, community-related, and location measures, summarized in Tables 1 and 2. We compare our data set with national statistics on the disease condition. National statistics indicate that individuals who develop the disease are typically between the ages of 40 to 70, with a median age of 55. Our sample has a median age of 56. Disease incidence in the United States is 60 percent and 40 percent for men and women respectively; our sample contains 57 percent men and 43 percent women. We note that the number of years suffering from the rare disease exhibits considerable variance, and multiple onset areas are represented in the sample. Patient tenure in the community ranges from one to six years, with a median of three. Overall, we have a sample of patients whose profile is consistent with the general population diagnosed with the disease, and there is sufficient variation in disease-related characteristics, with respect to time since diagnosis and the specific onset area of the disease, to support meaningful statistical analysis.

In addition to patient-level disease and demographic data, we collected all messages posted on the forum during the study period. The forum contains multiple threads, where a thread consists of an initiating topic or first post on the forum followed by subsequent replies. We use the terminology *originating post* to refer to the first post, *response* for subsequent posts or replies, and *thread* to refer to the originating post and its responses. After eliminating messages that were not posted by patients or those posted by patients with no location information or based outside the United States, our community interaction data consists of 4,247 threads, which contain 43,361 posts.

We developed a directed unweighted graph using community forum posts, to signify the observed support network of the online health community in statnet (Handcock et al. 2008). We used the following restrictions for constructing the graph: Each node in the network represents a patient who participated in the forum. There is directionality associated with support provision such that a supportive tie between a patient who posts a thread and a response from another patient is represented by a directed dyadic tie (unweighted), where the arrow points toward the originating poster and the arrow head terminating at the recipient (e.g., a patient whose initial post generates a reply in the thread would have a tie that is directed toward her). No loops are allowed in the graph, and the graph is binary as all ties are flattened such that a tie between A and B indicates at least one interaction occurred between the two nodes. Flattening ties is a conservative approach to graph construction as it results in a downward bias of the estimates of the models. Since our interest is in the location of the users who receive or provide information, using a binary graph does not affect the interpretation of the flow of information (Ackland and O'Neil 2011). Summary statistics of various indicators of centrality for the support network are provided in Table 3.

Statistical Analysis: Exponential Random Graph Models

We analyzed network interaction patterns to quantify social support exchanges using ERGM. ERGM is a stochastic network modeling method for deriving the likelihood of a net work emerging from all the possible structures that could have

[4]For simplicity, metropolitan areas are defined as *urban* and nonmetropolitan areas are defined as *rural*. The term *metropolitan area* is used to refer to an area with a population core of at least 2.5 million.

Table 1. Summary Statistics						
Variable	**Obs**	**Median**	**Mean**	**S.D.**	**Min**	**Max**
Demographic Covariates						
Age (year)	638	56	55.33	11.04	16	85
Location (1 = urban, 0 = rural)	638	1	0.83	0.38	0	1
Disease						
Years suffering from the disease	638	5	6.26	4.99	1	43
Community						
Membership (year)	638	3	3.56	1.09	1	6

Table 2. Frequency Distribution		
Variable	**Frequency**	**Percentage**
Location		
Rural	111	17.40
Urban	527	82.60
Gender		
Female	274	42.95
Male	364	57.05
Onset area		
Arms	239	37.46
Bulbar	160	25.08
Legs	212	33.23
Respiratory	27	4.23

Table 3. Summary Statistics for Centrality For Binary Network							
Variable	**Description**	**Obs**	**Median**	**Mean**	**S.D.**	**Min**	**Max**
Degree centrality	Number of incoming and outgoing ties for node	638	6.00	22.18	41.98	0	410
Indegree	Number of incoming ties for node	638	3.00	11.09	22.35	0	229
Rural indegree	Number of incoming ties for rural nodes	111	2.00	11.06	26.93	0	229
Urban indegree	Number of incoming ties for urban nodes	527	3.00	11.10	21.29	0	170
Outdegree	Number of outgoing ties for node	638	2.00	11.09	22.09	0	181
Rural outdegree	Number of outgoing ties for rural nodes	111	2.00	10.93	22.41	0	181
Urban outdegree	Number of outgoing ties for urban nodes	527	2.00	11.13	22.04	0	143

been formed by a random assignment of ties across nodes in the network. The model class is specified as

$$P(Y = y|X) = \exp\{(\theta^T g(y,X))\}/k(\theta)$$

where Y is the random set of relations in a network, y is a specific set of relations, X is a matrix of nodal attributes, g(y,X) is a vector of network statistics, θ is a vector of coefficients, and k(θ) is a normalizing constant that ensures that the sum of probability equals one.[5] The probability of graph y appearing across the distribution of graphs Y|X and k(θ) is

[5]Any network can be expressed as the exponential function of the counts of network statistics g(y,X). For a specific network, the probability of this network being generated is the exponential function of a set of network statistics for that particular network divided by the sum of exponential functions of the same set of network statistics across all possible networks in the distribution of the graph.

Table 4. Advantages of ERGM (Lusher et al. 2013)

Advantage	Description
Dependent ties in network	ERGMs allow for dependence among ties. In contrast, most traditional network modeling techniques like regression assume that network ties are independent. This assumption is unrealistic as it has been well established in previous studies that prior ties in a network will influence the formation of future ties. Researchers working on studies related to online social networks will find this relaxed assumption to be closer to their observed network.
Self-organizing property	ERGMs model social networks ties using a bottom up approach which is consistent with the widely observed self-organizing property of social networks. For instance, in an online patient network, patients often do not know each other until they start to post messages on the forums. The ties are gradually built up over time.
Multiple and simultaneous theories	ERGMs allow researchers to include multiple network dynamics and processes simultaneously. In other words, this technique provides researchers the capability to incorporate more than one theory into the same model at the same time. As an example, previous studies have established that network dynamics such as homophily and reciprocity occur in a network simultaneously. This approach gives researchers the flexibility to include various parameters in the network such that their models are as close as possible to the observed networks.
Multilevel attributes	A major advantage of ERGMs as compared to the traditional regression techniques in modeling social networks is the ability to incorporate both actors' attributes, and dyadic and network attributes into the model. A typical example will be ERGM's ability to incorporate individual attributes (e.g., patient's gender) while considering reciprocity in the same model.

estimated using MCMC. From these two estimates, the θ coefficient can be mathematically derived. The general interpretation of the significance and direction of the coefficients and the coefficients can be transformed as odds ratios using an exponential transformation (e^{θ}). While not equivalent to logistic regression, the formula can be reexpressed as the log odds that any given edge will exist, conditional on the current state of the rest of the network

$$\text{logit}(Y_{ij} = 1) = \theta^{T}\text{delta}[g(y,X)]_{ij}/k(\theta)$$

where Y_{ij} is an actor pair in Y and delta$[g(y,X)]_{ij}$ is the change in g(y,X) when the value of Y_{ij} is changed from 0 to 1. ERGM is a multivariate approach, which allows us to isolate the effect of location while controlling for other confounding factors.

As noted before, the traditional approach of examining the impact of location on the formation of a tie between any two nodes using methods such as logistic regression imposes strong statistical assumptions that most social network data violate. In particular, a key assumption underlying such methods that is required in order to compute standard errors is that ties are independent observations from each other. This assumption is overly limiting due to the nature of network data, because higher structural dependencies such as triadic closure have been shown to influence relationships between nodes in a network (Goodreau et al. 2009).

Compared to the traditional approach, ERGM offers several advantages to model the network, as summarized in Table 4. Importantly, it allows for dependence among ties, and mimics the self-organizing process of a social network. Multiple network dynamics such as homophily and reciprocity can be incorporated simultaneously as well. ERGM analysis allows us to include both patient and network level measures. Each node (i.e., patient) within the network has a number of specific demographic and disease related characteristics associated with it, while the community as a whole generates a range of network level measures that are included in the analysis. More detailed overviews of ERGM are available in Harris (2013), Handcock et al. (2008), Lusher et al. (2013), Robins and Morris (2007), and Robins et al. (2007).

Measures

Table 5 describes the set of network statistics used to generate the family of graphs for the application of ERGM to our data. The interpretation of these variables is discussed below.

Location Measures

The primary independent variable of interest is the location of the patient and the exchange of social support, conditional on location. We use two approaches to capture the effect of

Table 5. Network Variables

Network Statistic	Network Statistic Definition and Variable Construction
Edges	A measure of density of the network. The network statistic counts the number of ties, also referred to as edges, in the network.
AbsDiff(*x*)	A measure of the tendency of tie formation between nodes due to the difference of attribute *x*. The corresponding network statistic sums the absolute distance (absolute difference or AbsDiff) between every edge in the network. Specifically, we included *AbsDiff*(*years suffering from rare disease*) to control for the tendency of active experienced patients to interact; *AbsDiff(age)* to control for the tendency for patients with similar age to interact and *AbsDiff(membership)* to control for the impact of members that joined at different periods and the extent to which they interact with one another.
NodeMatch(*x*)	A measure of the degree of similarity between the two nodes in a tie for a specified attribute *x*. A positive coefficient suggests that patients with the same attribute *x* are more likely to form a tie. In our model, we include *NodeMatch*(*gender*) to control for the impact of gender similarities between dyads and *NodeMatch*(*onset of disease*) to control for the propensity of tie to form among patients with similar onsets, i.e. the area of the body where the patient first experienced symptoms of the disease.
NodeInFactor(*x*)	A measure of the degree to which a node with attribute *x* has the propensity to form incoming ties (i.e. establish connections coming in from other nodes). A positive coefficient of this variable indicates a higher chance of a node with attribute *x* to form incoming ties than a node without attribute *x,* other things equal. In our model, *NodeInFactor(rural)* is added to measure the propensity of rural nodes to form incoming ties as compared to urban nodes. In the empirical estimation, the corresponding network statistic is calculated as the number of times a node with a specified attribute *x* appears in incoming ties (or edges) for a recipient node in the network. For example, for a network consisting of one rural node with three incoming ties, the corresponding network statistic of *NodeInFactor(rural)* will be 3. If the network has three rural nodes, two with one incoming tie each, and one with three incoming ties, the corresponding network statistic of *NodeInFactor(rural)* will be 5.
NodeOutFactor(*x*)	A measure of the degree to which nodes with attribute *x* have the propensity to form outgoing ties with other nodes in the network. In our model, *NodeOutFactor(rural)* is added to measure the propensity of rural nodes to form outgoing ties.
Reciprocity (Mutual)	A measure of reciprocity in the network, which is estimated from a network statistic equal to the number of pairs of nodes *i* and *j* for which edges in both directions exist.
Triadic Closure (GWESP)	This measure captures the tendency for a patient to be attracted to other patients who have previously interacted with each other. To estimate this tendency to form closed triangles in the network, we construct the triadic closure using geometrically weighted edgewise shared partner distribution (*GWESP*). A parametric form of this count distribution gives each extra shared partner a declining impact on the probability of forming a tie (Goodreau et al. 2008; Hunter, Handcock et al. 2008). This metric has been shown to work well in avoiding model degeneracy, which arises when using Markov Chain Monte Carlo based estimation for graph models, and can result in estimations that are not consistent or graphs that are either empty or complete.

location on social support. First, we examine if the likelihood of forming a tie within the network is different for urban users than for rural ones. The difference, if it exists, will be reflected in the following two variables in the model: *NodeInFactor(rural)* and *NodeOutFactor(rural)*. To illustrate, a positive coefficient for the *NodeInFactor(rural)* suggests that if a node is in a rural location, it has higher marginal odds of receiving responses, compared to an urban node (Ackland and O'Neil 2011). If the coefficient is negative, it implies that the rural node has a smaller likelihood of receiving social support compared to an urban node.[6] By contrast, a positive coefficient for the second measure,

[6] *NodeInFactor(rural)* is a covariate in the model. While ERGM does not reduce to logistic regression, the coefficient for *NodeInFactor(rural)* can be interpreted as the conditional log odds of a tie received by a rural node compared to that of an urban node. Thus, a positive coefficient reflects a higher probability for a tie to appear as an incoming tie to a rural node, whereas a negative coefficient indicates lower probability for a tie to appear as an incoming tie to a rural node.

NodeOutFactor(rural), implies that rural nodes are more likely to provide support relative to urban nodes. If our assertion that urban patients can help alleviate the social support deficit for rural patients is true, the coefficient for *NodeInFactor(rural)* should be positive, indicating that rural nodes are more likely to be recipients, and *NodeOutFactor(rural)* should have a negative coefficient, indicating that urban nodes are more likely to be providers. Our second approach to measure the location effect is to examine the formation of directed ties. The directed ties can be divided into four types: *Rural* → *Rural*, *Urban* → *Rural*, *Rural* → *Urban*, and *Urban* → *Urban*. *Urban* → *Rural* means that a rural node has an incoming edge from an urban node, and so on. To avoid perfect multicollinearity, we use *Urban* → *Urban* as the base case in the statistical models. The coefficients of these variables are interpreted as the likelihood of forming the corresponding type of tie relative to an *Urban* → *Urban* tie.

Other Network Variables

To eliminate variance in tie formation and network exchange patterns that could plausibly be generated by other patient characteristics, we include a robust set of controls in the empirical specification. The choice of controls is based on prior work in tie formation that has found a positive association between homophily[7] (McPherson et al. 2001; Preece 2000) and reciprocity (Baker and Bulkley 2014; Johnson et al. 2014) on the propensity for a network tie to form. To account for the length of time different users may have had an opportunity to interact with each other, we add a control for community tenure.

Findings

Main Results

Using ERGM we are able to test if the observed support network is random or if it is an outcome of exchange patterns proposed here (i.e., after controlling for other factors that may affect tie formation, there is a net surplus of social support ties from urban to rural patients). Table 6 presents estimation results from a series of ERGM models. Model 1 is the baseline model, with a constant edge term, equivalent to network density that provides the baseline probability of one node forming a tie with another node in the network.

[7] *Homophily* refers to the phenomenon in which similarity on individual attributes such as age, religion, and education increases the likelihood of a relationship or a tie between two individuals.

In Model 2 we include the covariates for disease, demographics, and community. For the two covariates related to the disease, one measures the absolute difference between the number of years the patients have suffered from the disease and the second measures the similarity between two nodes with respect to disease onset. We also introduce higher-order terms for triadic closure (*GWESP*) and reciprocity (*Mutual*) as controls. These terms help to determine whether the observed tendency for homophilic sender and receiver ties is amplified by the balancing mechanisms of reciprocation and closed triangles, independent of the characteristics of the latter. A positive and significant coefficient for reciprocity implies that if patient A responds to patient B, there is tendency for patient B to respond back to patient A. A positive and significant coefficient for triadic closure implies that if patients D and E respond to patient A, there is a tendency for D and E to interact with each other, forming a triangle between three patients.

Results show that both disease covariates are significant. The positive coefficient for *AbsDiff(years suffering from rare disease)* ($p < 0.001$) indicates that the greater the difference between the nodes, the higher the propensity for a tie between them. The positive coefficient for the variable *NodeMatch(onset of disease)* suggests that patients with similar disease onset experiences are more likely to form a tie. For example, a patient whose disease onset started with the arms is more likely to communicate with another patient with the same onset region. Again, this is expected as patients are likely to have a predisposition to seek out patients with similar disease characteristics, since those individuals would be the most capable of providing relevant and useful responses.

The community level control of tenure, *AbsDiff(membership)*, captures the difference between two connected nodes based on the number of years elapsed since joining the community. Our results show that the coefficient of the difference in tenure of patients in the community is significant ($p < 0.001$) and has a negative effect on the propensity of tie formation, conditional on other network statistics. This finding is consistent with more generalized studies of social groups and online communities where tenure often has been found to play an important role in predicting individual contribution behavior (Ancona and Caldwell 1992; Kraut et al. 2002). Our results show that individuals connect with others who became active in the community contemporaneously: the smaller the difference in tenure between the two nodes, the more likely that a supportive tie between the dyad will form.

In Model 3, the two critical measures of social value creation related to location, *NodeInFactor(rural)* and *NodeOutFactor(rural)*, are added. From the results for Model

Table 6. Exponential Random Graph Models				
Covariates	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Edges	-4.096 <1e-04 ***	-4.887 <1e-04 ***	-5.254 <1e-04 ***	-5.229 <1e-04 ***
Disease				
AbsDiff(years suffering from rare disease)		0.019 0.0005***	0.0172 <1e-04 ***	0.0170 <1e-04 ***
NodeMatch(onset of disease)		0.098 <1e-04 ***	0.085 <1e-04 ***	0.087 <1e-04 ***
Demographic				
AbsDiff(age)		-0.0067 0.667	-0.006 <1e-04 ***	-0.004 0.530
NodeMatch(gender)		-0.090 0.660	0.0347 0.181	-0.031 0.647
Community				
AbsDiff(membership)		-0.260 <1e-04 ***	-0.256 <1e-04 ***	-0.257 <1e-04 ***
Balancing mechanisms				
Reciprocity (*Mutual*)		2.578 <1e-04 ***	2.309 <1e-04 ***	2.311 <1e-04 ***
Triadic Closure (*GWESP*)		1.410 <1e-04 ***	1.534 <1e-04 ***	1.534 <1e-04 ***
Location effects				
NodeInFactor(rural)			0.0672 <1e-04 ***	
NodeOutFactor(rural)			-0.0166 <1e-04 ***	
Rural → Rural				0.049 0.714
Urban → Rural				0.051 <1e-04 ***
Rural → Urban				-0.031 <1e-04 ***
AIC	67711	58498	53349	53343
BIC	67722	58585	53469	53463
Number of patients	638	638	638	638

***Significant at 1 percent level. **Significant at 5 percent level.

3, we find that *NodeInFactor(rural)* is positive and significant (0.0672, $p < 0.001$). This means that the probability of a node with an incoming tie is 7 percent higher for a rural node as compared to an urban node. In other words, all else equal, a rural node is more likely to receive support from other nodes. We also find that the coefficient for *NodeOutFactor(rural)* is negative and significant (-0.0166, $p < 0.001$), implying that rural patients are less likely than their urban counterparts to provide support, and the magnitude of the coefficient suggests that the probability of a rural node with an outgoing tie is 2 percent lower than an urban node. Taken together, these results provide support for the claim that there is a net surplus of social support flowing from urban to rural users. The coefficients for the higher order structural variables *Reciprocity (Mutual)* and *Triadic Closure (GWESP)* are positive and significant (2.309 and 1.534, respectively, $p < 0.001$). The direction and significance of other variables is similar to those in Model 2.

While the variables *NodeInFactor(rural)* and *NodeOutFactor(rural)* measure how connections are made to and from a rural node in Model 3, they do not reflect where the support flows to or from with respect to node location. In Model 4, we use directed network measures which estimate the likelihood of support flow occurring across locations *Urban* → *Rural*, *Rural* → *Rural*, and *Rural* → *Urban* to examine the structural patterns of supportive ties between urban and rural patients. The excluded category is *Urban* → *Urban*. The coefficient for *Urban* → *Rural* is positive and significant (0.051, $p < 0.001$). This suggests that an *Urban* → *Rural* tie has a greater probability of existence than the baseline case, which is *Urban* → *Urban*. In other words, the likelihood of an urban patient responding to a rural patient is higher than the likelihood of responding to another urban patient, all else equal. Furthermore, the coefficient for *Rural* → *Urban* is significant and negative (-0.031, $p < 0.001$), indicating that this structural pattern of rural patients providing support to urban patients occurs less frequently than in randomly generated graphs and there is a decreased propensity for supportive tie formation. Interestingly, we find that a *Rural* → *Rural* tie is not significant, indicating that this type of tie is as likely to form as the default *Urban* → *Urban* tie. Among the control variables, consistent with previous work on online social networks (Szell and Thurner 2010; Wimmer and Lewis 2010) *Reciprocity* (*Mutual)* and *Triadic Closure* (*GWESP*) remain positive and significant in this model.

Goodness of Fit for ERGM Models

We estimated the goodness of fit for the ERGM models in two ways. First, we calculated Akaike's Information Criterion or AIC (Akaike 1987) and Bayes Information Criterion or BIC (Schwarz 1978) values. Inspection of the differences in the information criterion values for Models 1–4 in Table 6 reveals that Models 3 and 4 exhibit better fit than the rest of the models.

Second, we used the commonly deployed approach whereby simulations of the proposed model are compared to the actual network (Goodreau et al. 2008; Hunter, Goodreau, and Handcock 2008) and examined the properties that were not explicitly modeled. For a given network statistic (such as the indegree), if the simulated model is close to that of the actual network, this implies a greater degree of confidence in the model. We generated 1,000 simulations of random networks for each model and examined where the real network statistics occur relative to those of simulated networks. The simulations determine the extent to which microprocesses represented by a model are capable of reproducing key features of the network's global structure—specifically, the indegree, outdegree, edge-wise, dyad-wise shared partners, and geodesic distances. If the models accurately represent observed information exchanges, the distribution for these two networks should be similar.

Simulation results[8] for the dyadic dependent ERGMs produced a goodness of fit plot that fit the original data well. The model appears to have slightly under-predicted edge-wise shared partners for small values and over-predicted for larger values. The other graph level indices were predicted reasonably well.

Robustness Tests

We conducted an extensive series of robustness tests to eliminate alternative explanations for the findings, as summarized in Table 7. First, it could be argued that the findings might be driven by message exchanges that are of marginal value to rural patients and health capacity. To address this concern, we limited our analyses to two types of posts respectively: posts starting with question marks, and posts that are marked as "helpful." Results were consistent with the main findings. We further conducted analyses using subcategories of support (informational, emotional, and experiential) and found that rural patients receive net surplus in all three types of support. This confirms that our findings are not sensitive to the specific types of social support.

Second, to determine if the findings are sensitive to the specific network dynamics and structure, such as where some individuals contribute a disproportionate share of posts, we performed ERGM analyses by eliminating top contributors (specifically, the top 10, 15, and 20) from the sample to address the confound. Results consistently show that urban patients are more likely to contribute to rural patients in terms of social support. Next, we performed simulation tests by varying the number urban nodes (ranging from 100 to 500), while keeping the rural patient nodes constant, to confirm that our findings are not driven by the fact that there are more urban patients than rural ones. Finally, we found no statistical differences in the means of age, disease, or community covariates between urban and rural patients.[9]

---

[8]Details of these additional tests are not shown due to space limitations but available on request.

[9]To confirm if there is indeed a disparity in access to specialists, we extracted only the specialists for this type of rare disease and the population information for all zip codes in the United States from the ARF. We find a significant difference between the specialist to population ratio for rural and urban areas. This supports previous findings that urban residents have better access to physicians than rural residents (Rosenthal et al. 2005).

Table 7. Robustness Checks[†]

	Are the results the same when we account for different message characteristics?				
	Messages with question marks	*Messages rated helpful*	*Messages with informational support*	*Messages with emotional support*	*Messages with experiential support*
Urban → Rural	✓ (+)***	✓ (+)***	✓ (+)**	✓ (+)**	✓ (+)**
Rural → Urban	✓ (-)***	× (-)	× (+)	✓ (-)**	✓ (-)***

	Are the results the same when we account for network characteristics?							
	Top 10 contributors removed	*Top 15 contributors removed*	*Top 20 contributors removed*	*100 urban users*	*200 urban users*	*300 urban users*	*400 urban users*	*500 urban users*
Urban → Rural	✓ (+)***	✓ (+)***	✓ (+)***	✓ (+)***	✓ (+)**	✓ (+)***	✓ (+)***	✓ (+)***
Rural → Urban	✓ (-)***	✓ (-)***	✓ (-)***	✓ (-)***	✓ (-)***	✓ (-)***	✓ (-)***	✓ (-)***

[†]Details of these robustness checks are not shown due to space limitations but are available upon request.
✓ Found statistical significant evidence at 5% level.
× Not significant at 5 percent level .
+/- Indicates the sign of the coefficient for variable indicated in row (i.e., Urban → Rural or Rural → Urban).
***Significant at 1 percent level.
**Significant at 5 percent level.

Overall, results from the main ERGM analysis, consistent across robustness checks, indicate the existence of a net surplus of social support from urban to rural patients. Based on these findings, we conclude that participation in online health communities enables the amplification of rural patients' health capabilities. To the degree that improvement in health capabilities is an indicator of a reduction of health disparities, we find evidence for our assertion that online health communities generate social value.

Discussion and Conclusion

The accelerating growth of online health communities and robust participation by patients have created an opportunity for these networks to positively influence public health. Previous studies have pointed to the unmet needs of patients as a result of their geographical location and documented the presence of significant health disparities between rural and urban residents. We argued that a leading cause of such disparities is the existence of health capability gaps among rural residents, and that online communities can reduce these gaps, thereby generating social value. To isolate social value, we modeled the online social support of patients participating in an online health community where nodes in a network represented the patients who are interconnected on multiple dimensions. Using data collected from a popular online rare disease community, we found empirical evidence to support the assertion that urban patients are net suppliers of online social support to rural patients.

This study makes two important contributions to the literature. One, to the best of our knowledge, this is the first study focusing on social value creation in online communities. Two, our use of ERGM as the empirical method allows us to analyze the network while controlling for structural characteristics, thereby overcoming a major limitation of traditional techniques. While it is generally acknowledged that a network approach is useful, the specific mechanisms and the relationships between different types of units in social networks remain far less understood (Jackson 2010). There is a need to study the relations between different types of units to develop deeper insights into social networks. By viewing participants as interacting agents that give rise to different structures and patterns within networks, new questions about the specific forms of interactions and interdependence can be answered. This approach could also be used to improve the explanatory potential of future research, especially those proposing middle range theories (McGrath 2013). Previous research has lacked the ability to demonstrate quantitatively the presence of generative mechanisms and account for endogeneity (Johnson et al. 2014); this study illustrates how ERGM can be used to address these issues.

We demonstrated how to model information flows and supportive ties between different types of sources and recipients in online health communities as a network using ERGM, thereby providing a foundation for future research in the information systems discipline that is increasingly turning its attention to social networks. Researchers in information systems can use ERGM to explore multiple theories simultaneously as well as take endogenous factors (e.g., individual attributes) into account, which was previously not possible. For instance, in extending the current study, researchers can examine the value of online health communities for other stigmatized groups by controlling for both individual and network level attributes. Our use of the methodology should generate further interest and facilitate its application in a multitude of fields.

We acknowledge the limitations of the study that present useful opportunities for future work. First, the demographic data on the site is self-reported and therefore may suffer from self-reporting biases. However, because participation in the community is volitional and individuals join because of a sense of affiliation and shared identity with fellow disease sufferers, we have less reason to suspect that individuals would provide false data. Second, as the data is collected from a single community focusing on a rare disease, the generalizability of the findings to other types of online health communities should be viewed with caution. In addition, while we have included extensive controls for patient heterogeneity in the analysis, there still may be unobserved factors associated with patients that might bias our findings. Third, our ERGM models were modeled using a directional binary network. Although a binary representation provides more conservative findings, future work may consider using weighted networks when tools in this area become more developed. Fourth, this study focuses on one particular form of social value creation: bridging the regional health knowledge gap by enabling knowledge flow from urban to rural patients. The online community enables other forms of social value creation as well; for example, any patient who browses the threads will likely benefit from the online community, even if they do not post a message. Therefore, the evidence we presented in this paper should be interpreted as a conservative estimate reflecting a subset of the total social value of the online community. Finally, future research can also examine the actual health benefits of such online social support for patients, which the current study does not do.

Our findings yield implications for policy makers and practitioners concerned with meeting patient needs and overcoming disparities in medical access. Entities responsible for resource allocation decisions, such as governments, community agencies, and public health facilities, should leverage the powerful role that online collectives can play. Online communities can serve as a low cost alternative to or as a complement to existing health programs. This study also demonstrates the potential of using online social media data to drive evidence-based decision making. The freely available online data represents a new channel to uncover the latent needs of disadvantaged patients. In this regard, our results underscore the importance of policy initiatives related to expanding Internet access and broadband infrastructure in rural areas, as patients residing in these areas can benefit more by exchanging information with the outside world. Globally, the relative simplicity and low cost of online communities implies that they could be replicated quickly and yield a positive impact in nations where residents of remote areas are often disadvantaged with respect to limited access to health services. Finally, it is important to note that the social media data can generate significantly more insights and practical value when combined with patient clinical data and other behavioral measures. Advanced research data infrastructure that provides an integrated view of all aspects of patient activities could significantly facilitate our exploitation of the value generated from online patient communities.

Acknowledgments

The first author gratefully acknowledges the financial support of the Fundación Ramón Areces and IE Business School.

References

Ackland, R., and O'Neil, M. 2011. "Online Collective Identity: The Case of the Environmental Movement," *Social Networks* (33:3), pp. 177-190.

Adler, N. E., and Rehkopf, D. H. 2008. "U.S. Disparities in Health: Descriptions, Causes, and Mechanisms," *Annual Review of Public Health* (29:1), pp. 235-252.

Ahern, K., and Harford, J. 2014. "The Importance of Industry Links in Merger Waves," *Journal of Finance* (69:2), pp. 527-576.

Akaike, H. 1987. "Factor Analysis and AIC," *Psychometrika* (52:3), pp. 317-332.

Ancona, D. G., and Caldwell, D. F. 1992. "Demography and Design: Predictors of New Product Team Performance," *Organization Science* (3:3), pp. 321-341.

Armstrong, A., and Hagel, J. 2000. "The Real Value of Online Communities," in *Knowledge and Communities*, E. Lesser, M. Fontaine, and J. Slusher (eds.), Oxford, UK: Butterworth-Heinemann, pp. 85-95.

Baker, W. E., and Bulkley, N. 2014. "Paying it Forward vs. Rewarding Reputation: Mechanisms of Generalized Reciprocity," *Organization Science* (25:5), pp. 1493-1510.

Baldwin, L.-M., MacLehose, R. F., Hart, L. G., Beaver, S. K., Every, N., and Chan, L. 2004. "Quality of Care for Acute Myocardial Infarction in Rural and Urban US Hospitals," *The Journal of Rural Health* (20:2), pp. 99-108.

Bateman, P. J., Gray, P. H., and Butler, B. S. 2011. "The Impact of Community Commitment on Participation in Online Communities," *Information Systems Research* (22:4), pp. 841-854.

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., and Crotty, K. 2011. "Low Health Literacy and Health Outcomes: An Updated Systematic Review," *Annals of Internal Medicine* (155:2), pp. 97-107.

Bunde, M., Suls, J., Martin, R., and Barnett, K. 2006. "Hystersisters Online: Social Support and Social Comparison Among Hysterectomy Patients on the Internet," *Annals of Behavioral Medicine* (31:3), pp. 271-278.

Cranmer, S. J., and Desmarais, B. A. 2011. "Inferential Network Analysis with Exponential Random Graph Models," *Political Analysis* (19:1), pp. 66-86.

Edwards, L. 2013. *In the Kingdom of the Sick*, New York: Walker Publishing Company, Inc.

Faraj, S., and Johnson, S. L. 2011. "Network Exchange Patterns in Online Communities," *Organization Science* (22:6), pp. 1464-1480.

Faraj, S., Kudaravalli, S., and Wasko, M. 2015. "Leading Collaboration In Online Communities," *MIS Quarterly* (39:2), pp. 393-412.

Franke, N., and Shah, S. 2003. "How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users," *Research Policy* (32:1), pp. 157-178.

Gamm, L., Castillo, G., and Pittman, S. 2010. "Access to Quality Health Services in Rural Areas," *Primary Care* (1), pp. 45-75.

Goodreau, S. M., Handcock, M. S., Hunter, D. R., Butts, C. T., and Morris, M. 2008. "A Statnet Tutorial," *Journal of Statistical Software* (24:9), pp. 1-26.

Goodreau, S. M., Kitts, J. A., and Morris, M. 2009. "Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks," *Demography* (46:1), pp. 103-125.

Graydon, J., Galloway, S., Palmer-Wickham, S., Harrison, D., der Bij, L. R., West, P., Burlein-Hall, S., and den Evans-Boy, B. 1997. "Information Needs of Women During Early Treatment for Breast Cancer," *Journal of Advanced Nursing* (26:1), pp. 59-64.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. 2008. "statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data," *Journal of Statitical Software* (24:1).

Harris, J. K. 2013. *An Introduction to Exponential Random Graph Modeling*, Thousand Oaks, CA: Sage Publications.

House, J. S. 2002. "Understanding Social Factors and Inequalities in Health: 20th Century Progress and 21st Century Prospects," *Journal of Health and Social Behavior* (43:2), pp. 125-142.

Huang, Y., Singh, P. V., and Srinivasan, K. 2014. "Crowdsourcing New Product Ideas Under Consumer Learning," *Management Science* (60:9), pp. 2138-2159.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. 2008. "Goodness of Fit of Social Network Models," *Journal of the American Statistical Association* (103:481), pp. 248-258.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. 2008. "ERGM: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks," *Journal of Statistical Software* (24:3), p. nihpa54860.

Institute of Medicine of the National Academies. 2005. *Quality Through Collaboration: The Future of Rural Health*, Washington DC: National Academies Press.

Jackson, M. O. 2010. *Social and Economic Networks*, Princeton, NJ: Princeton University Press.

Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. 2012. "Social Capital and Social Quilts: Network Patterns of Favor Exchange," *The American Economic Review* (102:5), pp. 1857-1897.

Johnson, G. J., and Ambrose, P. J. 2006. "Neo-Tribes: The Power and Potential of Online Communities in Health Care," *Communications of ACM* (49:1), pp. 107-113.

Johnson, S. L., Faraj, S., and Kudaravalli, S. 2014. "Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment," *MIS Quarterly* (38:3), pp. 795-808.

Kane, G. C., and Fichman, R. G. 2009. "The Shoemaker's Children: Using Wikis for Information Systems Teaching, Research, and Publication," *MIS Quarterly* (33:1), pp. 1-17.

Koh, H. K., Graham, G., and Glied, S. A. 2011. "Reducing Racial and Ethnic Disparities: The Action Plan from the Department of Health and Human Services," *Health Affairs* (30:10), pp. 1822-1829.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., and Crawford, A. 2002. "Internet Paradox Revisited," *Journal of Social Issues* (58:1), pp. 49-74.

Lengerich, E. J., Tucker, T. C., Powell, R. K., Colsher, P., Lehman, E., Ward, A. J., Siedlecki, J. C., and Wyatt, S. W. 2005. "Cancer Incidence in Kentucky, Pennsylvania, and West Virginia: Disparities in Appalachia," *Journal of Rural Health* (21:1), pp. 39-47.

Lipsky, M. S., and Glasser, M. 2011. "Critical Access Hospitals and the Challenges to Quality Care," *Journal of the American Medical Association* (306:1), pp. 96-97.

Lusher, D., Koskinen, J., and Robins, G. (eds.). 2012. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*, Cambridge, UK: Cambridge University Press.

Ma, M., and Agarwal, R. 2007. "Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities," *Information Systems Research* (18:1), pp. 42-67.

Manchanda, P., Packard, G., and Pattabhiramaiah, A. 2015. "Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Community," *Marketing Science* (34:3), pp. 367-387.

Martin, A. B., Torres, M., Vyavaharkar, M., Chen, Z., Towne, S., and Probst, J. C. 2013. *Rural Border Health Chartbook*, Columbia, SC: South Carolina Rural Health Research Center.

McGrath, K. 2013. ¨The Potential of Generative Mechanisms for IS Research," in *Proceedings of the 34th International Conference on Information Systems*, Milan, Italy, December 15-18.

McPherson, M., Smith-Lovin, L., and Cook, J. M. 2001. "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology* (27:1), pp. 415-444.

Nevo, D., and Furneaux, B. 2012. "The Power of Communities: From Observed Outcomes to Measurable Performance," in *Proceedings of the 33rd International Conference on Information Systems*, Orlando, FL.

Nobles, J., and Frankenberg, E. 2009. "Mothers' Community Participation and Child Health," *Journal of Health and Social Behavior* (50:1), pp. 16-30.

O'Grady, M. J., Mueller, C., and Wilensky, G. R. 2002. "Essential Research Issues in Rural Health: The State Rural Health Director's Perspective," *Policy Analysis Brief, Series W* (5:1), pp. 1-4.

Porter, C. E., and Donthu, N. 2008. "Cultivating Trust and Harvesting Value in Virtual Communities," *Management Science* (54:1), pp. 113-128.

Preece, J. 2000. *Online Communities: Designing Usability and Supporting Socialbility* (1st ed.) New York: John Wiley & Sons.

Ray, S., Kim, S. S., and Morris, J. G. 2014. "The Central Role of Engagement in Online Communities," *Information Systems Research* (25:3), pp. 528-546.

Ren, Y., Kraut, R., and Kiesler, S. 2007. "Applying Common Identity and Bond Theory to Design of Online Communities," *Organization Studies* (28:3), pp. 377-408.

Ricketts, T. C. 2000. "The Changing Nature of Rural Health Care," *Annual Review of Public Health* (21:1), pp. 639-657.

Robins, G., and Morris, M. 2007. "Advances in Exponential Random Graph Models," *Social Networks* (29:2), 169-172.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. 2007. "An Introduction to Exponential Random Graph (p*) Models for Social Networks," *Social Networks* (29:2), pp. 173-191.

Rosenthal, M. B., Zaslavsky, A., and Newhouse, J. P. 2005. "The Geographic Distribution of Physicians Revisited," *Health Services Research* (40:6p1), pp. 1931-1952.

Ruger, J. P. 2010. "Health Capability: Conceptualization and Operationalization," *American Journal of Public Health* (100:1), pp. 41-49.

Rutten, L. J., Hesse, B. W., Moser, R. P., Ortiz Martinez, A. P., Kornfeld, J., Vanderpool, R. C., Byrne, M., and Tortolero Luna, G. 2012. "Socioeconomic and Geographic Disparities in Health Information Seeking and Internet Use in Puerto Rico," *Journal of Medical Internet Research* (14:4).

Schur, C. L., and Franco, S. J. 1999. "Access to Health Care," in *Rural Health in the United States*, T. C. Ricketts (ed.), New York: Oxford University Press, pp. 25-37

Schwarz, G. 1978. "Estimating the Dimension of a Model," *The Annals of Statistics* (6:2), pp. 461-464.

Sen, A. 1999. *Development as Freedom*, New York: Alfred A. Knopf.

Sen, A. 2002. "Why Health Equity?," *Health Economics* (11:8), pp. 659-666.

Sun, Y., Fang, Y., and Lim, K. H. 2012. "Understanding Sustained Participation in Transactional Virtual Communities," *Decision Support Systems* (53:1), pp. 12-22.

Szell, M., and Thurner, S. 2010. "Measuring Social Dynamics in a Massive Multiplayer Online Game," *Social Networks* (32:4), pp. 313-329.

Tsai, H., and Bagozzi, R. P. 2014. "Contribution Behavior in Virtual Communities: Cognitive, Emotional, and Social Influences," *MIS Quarterly*, (38:1), pp. 143-163.

Trusov, M., Bucklin, R. E., and Pauwels, K. 2009. "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing* (73:5), pp. 90-102.

United Health. 2011. "Modernizing Rural Health Care: Coverage, Quality and Innovation," Working Paper 6, United Health Center for Health Reform & Modernization, Minnetonka, MN (http://www.unitedhealthgroup.com/hrm/UNH_WorkingPaper6.pdf).

Viswanath, K. 2006. "Public Communications and its Role in Reducing and Eliminating Health Disparities," *Examining the Health Disparities Research Plan of the National Institutes of Health: Unfinished Business*, G. E. Thomson, F. Mitchell, and M. B. Williams (eds.), Washington, DC: National Academies Press (http://www.ncbi.nlm.nih.gov/books/NBK57046/).

Wasko, M. M., and Faraj, S. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly* (29:1), pp. 35-57.

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., and Haythornthwaite, C. 1996. "Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community," *Annual Review of Sociology* (22), pp. 213-238.

Wimmer, A., and Lewis, K. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook," *The American Journal of Sociology* (116:2), pp. 583-642.

World Health Organization. 2010. *Increasing Access to Health Workers in Remote and Rural Areas Through Improved Retention: Global Policy Recommendations*, Geneva: World Health Organization.

Ziller E. C. 2014. "Access to Medical Care in Rural America," in *Rural Public Health: Best Practices and Preventive Models*, J. C. Warren and K. B. Smiley (eds.), New York: Springer Publishing Company, pp. 11-28.

About the Authors

Jie Mein (JM) Goh is an assistant professor of Management Information Systems in the Beedie School of Business at Simon Fraser University. Prior to joining Simon Fraser, she was a faculty at the IE Business School, IE University (Instituto de Empresa) in Madrid, Spain. JM obtained her Ph.D. at the Robert H Smith School of Business, University of Maryland, College Park, and her M.Sc. and B.Sc. (First Class Honours) from the School of Computing at the National University of Singapore. JM's current research focuses on the impact and assimilation of information technology in health care. Her work has been published in journals such as *Information Systems Research*, *MIS Quarterly*, *MIT Sloan Management Review*, *Journal of Strategic Information Systems*, among others.

Guodong (Gordon) Gao is an associate professor in the Decision, Operations, and Information Technology Department at the R. H. Smith School, University of Maryland, College Park, and codirector of the Center for Health Information and Decision Systems

(CHIDS). Gordon received his bachelor degrees from Tsinghua University, his MBA from the Tsinghua–MIT Sloan Joint Program, and his Ph.D. from the Wharton School of the University of Pennsylvania. His research focuses on the role of IT in transforming healthcare and improving quality transparency. His research has been published or presented at leading journals and conferences like *Management Science*, *Information Systems Research*, *MIS Quarterly*, *Manufacturing and Service Operations Management,* and *Production and Operations Management*. His work has also been published or presented in the healthcare outlets, such as *JAMA Internal Medicine*, *Journal of Medical Internet Research*, the International Health Economics Association Conference, as well as the American Society of Health Economists conference. He has won several teaching and research awards, including the CAREER award from the National Science Foundation.

Ritu Agarwal is a professor and the Robert H. Smith Dean's Chair of Information Systems at the Robert H. Smith School of Business, University of Maryland, College Park. She is also the founder and codirector of the Center for Health Information and Decision Systems (CHIDS) at the Smith School. She has published over 90 papers in journals such as *Information Systems Research*, *MIS Quarterly*, *Management Science*, *Communications of the ACM*, *Journal of Management Information Systems*, *JAMIA, Health Affairs, IEEE Transactions,* and *Decision Support Systems*. Her current research is focused at the intersection of information technology and healthcare, with a particular emphasis on mobile technologies, social media, digital information privacy, and evidence-based decision making. She is a Fellow of Association for Information Systems and the INFORMS Information Systems Society.